Data-driven optimization of seismicity models using diverse data sets: generation, evaluation and ranking using inlabru

Data-driven optimization of seismicity models using diverse data sets: generation, evaluation and ranking using inlabru

Kirsty Bayliss[1], Mark Naylor[1], Janine Illian[2], Ian G. Main [1]

[1]School of GeoSciences, University of Edinburgh, West Mains Road, Edinburgh
[2]School of Mathematics and Statistics, University of Glasgow, University Place, Glasgow

Key Points:

- Spatially varying seismicity can be efficiently modelled as a Log-Gaussian Cox process which includes deterministic and stochastic components
- LGCPs can be analysed with Integrated Nested Laplace Approximations to compare seismicity models and identify useful model components
- These models find maps of strain rate and distance to nearest fault useful for constraining spatial seismicity

Corresponding author: Kirsty Bayliss, Kirsty.Bayliss@ed.ac.uk

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2020JB020226

Abstract

Recent developments in earthquake forecasting models have demonstrated the need for a robust method for identifying which model components are most beneficial to understanding spatial patterns of seismicity. Borrowing from ecology, we use Log-Gaussian Cox process models to describe the spatially varying intensity of earthquake locations. These models are constructed using elements which may influence earthquake locations, including the underlying fault map and past seismicity models, and a random field to account for any excess spatial variation that cannot be explained by deterministic model components. Comparing the alternative models allows the assessment of the performance of models of varying complexity composed of different components, and therefore identifies which elements are most useful for describing the distribution of earthquake locations. We demonstrate the effectiveness of this approach using synthetic data and by making use of the earthquake and fault information available for California, including an application to the 2019 Ridgecrest sequence. We show the flexibility of this modelling approach and how it might be applied in areas where we do not have the same abundance of detailed information. We find results consistent with existing literature on the performance of past seismicity models, that slip rates are beneficial for describing the spatial locations of larger magnitude events and that strain rate maps can constrain the spatial limits of seismicity in California. We also demonstrate that maps of distance to the nearest fault can benefit spatial models of seismicity, even those that also include the primary fault geometry used to construct them.

Plain Language Summary

Recently many statistical models for earthquake occurrence have been developed with the aim of improving earthquake forecasting. Several different underlying factors might control the location of earthquakes, but testing the significance of each of these factors with traditional approaches has not been straightforward and has restricted how well we can combine different successful model elements. We present a new approach using a point process model to map the spatial intensity of events. This method allows us to combine maps of factors which might affect the location of earthquakes with a random element that accounts for other spatial variation. This allows us to rapidly compare models with different components to see which are most helpful for describing the observed locations. We demonstrate this approach using synthetic data and real data from California as a whole and the 2019 Ridgecrest sequence in particular. Slip rates are found to be beneficial for explaining the spatial distribution of large magnitude events and strain rates are found useful for constraining spatial limits of observed seismicity. Constructing a fault-distance map can also improve models where many events cannot be directly linked to an individual fault.

1 Introduction

For a variety of reasons, including the lack of clear, reliable precursors, and the inherent non-linearity and complexity of the underlying process, the deterministic prediction of individual earthquakes remains an elusive goal (Jordan et al., 2011). Instead the focus has shifted to forecasting the probability of occurrence of a population of events in space and time, in an attempt to determine the degree of predictability of earthquakes (Vere-Jones, 1995; Field, 2007; Jordan & Jones, 2010). In order to make reliable forecasts, it is necessary to understand as much as possible about the spatio-temporal behaviour of earthquakes and the underlying processes that drive them.

Statistical point process models have been used to describe earthquake occurrence for many years (Vere-Jones & Davies, 1966; Vere-Jones, 1970; Ogata, 1998). The aim of these models is to describe the occurrence of earthquakes as a series of points in time, space, or both, with an appropriate 'mark' such as earthquake magnitude. With the cre-

ation of the Epidemic-Type Aftershock Sequence (ETAS) model and development of robust methods to estimate parameter values (Ogata, 1988; Ogata & Zhuang, 2006; Ogata, 2011; Veen & Schoenberg, 2008), point process models have been applied extensively in statistical seismology. For example the ETAS model is widely used for catalogue simulation and model testing (Helmstetter & Sornette, 2002a, 2002b, 2003; Helmstetter, 2003; Helmstetter et al., 2005; Nandan et al., 2017; Seif et al., 2017), as well as being used in aftershock forecasting models (Marzocchi et al., 2014; Taroni et al., 2018; Rhoades et al., 2018).

The uniform California earthquake rupture forecast (UCERF) is a fault-based model for forecasting seismicity in the state of California. The most recent implementations of this model (UCERF3) include a time-independent model that assumes the process is statistically stationary and memoryless (Field et al., 2014), a long-term time-dependent model incorporating memory of large past events (Field et al., 2015) and a short-term time-dependent forecast model (Field et al., 2017) which makes use of the ETAS model to forecast aftershock activity. These models are used in parallel in hazard assessment for the state of California for different applications, and consist of four main components: a fault model for the physical locations and architecture of known faults; slip rate models that estimate the slip and creep estimates for each individual fault from geodetic and geological data; event rate models which describe the long-term rate of earthquakes throughout the area; and a probability model that describes the likelihood of an earthquake occurring within a specified time period. The time-dependent models also include long-term renewal and short-term clustering processes, while the time-independent model is applied to probabilistic seismic hazard assessment (PSHA). Typically, different model configurations are selected by use of a logic tree, where each branch is given a weighting determined by expert judgement through workshops. This approach allows the construction of models containing different information which will then be included in a full rupture forecast model and allows the inclusion of uncertainties in model parameters. In the case of UCERF3, this results in a logic tree with a total of 5760 branches, requiring the use of high power computing facilities to produce the resulting forecast models.

The idea of hybrid forecasting models is not a new one. Marzocchi et al. (2012) suggest a Bayesian method for combining models and ranking them based on their respective Bayes factor. Rhoades et al. (2014) proposed multiplicative combinations of models from the RELM project to improve the forecasting ability of models compared to a smoothed seismicity model. Rhoades et al. (2015) expanded on this approach and applied it to events in New Zealand, in which they included a covariate that accounted for the distance to a mapped fault. In each of these examples there is a requirement for individual forecasts to be developed before combination and for the user to determine a weight for the individual model components, an issue which is highlighted particularly in Marzocchi et al. (2012). Nevertheless, there remains a significant component of epistemic uncertainty due to lack of data, notably on the occurrence rates of large magnitude events.

In this paper we address the question: how can we be sure which components of the model are most useful in describing the seismicity in a data-driven approach, i.e. without recourse to expert judgement? Seismology currently lacks both a straightforward method for the objective combination of useful earthquake data and a robust framework for the rapid evaluation of seismicity models with a full description of the associated uncertainty. Here we present a possible solution - the application of integrated nested Laplace approximations (INLA) for the spatial modelling of earthquake data by modelling seismicity as a log-Gaussian Cox process (LGCP). Once the most useful model components are identified these can be applied in a straightforward way to prospective forecasting models. We begin by outlining the theory underlying the INLA method, and justifying its use in a seismological context. We then demonstrate the ease with which models describing the longer-term spatial distribution of seismicity can be constructed and compared

using synthetic data and data used in the UCERF3 model. We highlight how INLA can be used to straightforwardly assess the contribution of different components of a model, how well each component can describe the observed spatial distribution of events, and how the inclusion of a random field in the analysis can help us identify what our model is lacking.

2 Theory

The earliest point process models for earthquakes were discussed by Vere-Jones (1970), who lamented that the mathematical methods required to evaluate these models did not exist at the time. Early models by Ogata (1988) demonstrated how point processes could be used for the temporal modelling of earthquake sequences, by combining the Poisson rate of independent or 'background' earthquakes and the Omori law for dependent aftershock events to account for temporal clustering. Though the ETAS model has seen many improvements over the years, the question of the most suitable spatial model for a spatio-temporal ETAS model still remains. An isotropic inverse power law distribution is often used (Ogata & Zhuang, 2006; Werner et al., 2011) though more general spatial kernels were proposed by (Ogata, 1998) and sophisticated alternatives that include fault data and Coulomb stress changes have also been suggested (Bach & Hainzl, 2012). By stacking global data, (Huc & Main, 2003) showed an inverse power law with an exponential tail was the most appropriate global average, implying a correlation length for triggering similar to the seismogenic thickness. An alternative point process approach would be to consider the spatial distribution of events first, and then extend such a model to be fully spatio-temporal. Given that the spatial intensity of earthquake occurrence is known to vary, and to be associated with underlying sub-surface conditions which cannot be directly observed, a method that allows a quantitative description of the stochastic nature of earthquakes in space is required. Here we propose to solve this problem by adopting a log-Gaussian Cox process (LGCP) which models a spatially-varying intensity process as a function of deterministic and stochastic effects. Such models can be rapidly constructed and evaluated using integrated nested Laplace approximations (INLA). Below, we outline the theory behind LGCPs and the INLA method for fitting them. INLA is a computationally-efficient way to construct models, so we can construct models with different combinations of potentially useful components and compare their performance with appropriate methods.

2.1 LGCPs and INLA

Log-Gaussian Cox processes are a popular class of model for spatial variability in ecology, as they allow some observed spatial pattern to be described by some deterministic location effects and a 'random field' component which describes any remaining spatial variability. Observations can then be modelled with a spatially-varying intensity function that describes a continuous stochastic process as a function of combined stochastic and deterministic effects (Diggle et al., 2013). Where current point process models for seismicity began by describing the temporal distribution of earthquakes, Log-Gaussian Cox processes aim to model the spatial variation of events using an inhomogeneous Poisson process, which can be extended to a fully spatio-temporal marked point process. In the case of earthquake data, our 'deterministic' model components will inevitably still contain uncertainty, and can range from observed data such as fault maps to spatial data models, such as smoothed seismicity or strain rate models.

The integrated nested Laplace approximation (INLA) approach is a computationally efficient alternative to Markov-chain Monte-Carlo (MCMC) approaches to Bayesian model fitting. The INLA method is incredibly flexible and has been widely applied to point process datasets in ecology (Illian et al., 2012; Sadykova et al., 2017; Dutra Silva et al., 2017; Yuan et al., 2017) and to some natural hazard examples including tornado

(Gómez-Rubio et al., 2015), wildfire (Díaz-Avalos et al., 2016) and landslide modelling (Lombardo et al., 2018).

2.1.1 Log-Gaussian Cox processes (LGCP)

Sometimes called the doubly stochastic Poisson process, the Cox process (Cox, 1955) is a generalisation of the Poisson process where the process intensity λ is itself stochastic in nature. Vere-Jones (1970) proposed using the Cox process to describe the spatial distribution of earthquakes, if clustering were removed. A log-Gaussian Cox process is a special case of the Cox process where the rate of events is determined by some underlying random field, which is assumed to be Gaussian in nature and vary spatially. The construction of these models allows the user to specify underlying factors that might describe the observed spatial distribution, for example how underlying soil characteristics might affect the location of a certain tree species (Illian et al., 2012). Multiple underlying factors can be included in the spatially-varying intensity function to account for the observed spatial distribution of events. A random field can then also be added, which describes the remaining spatial structure which cannot be captured by more deterministic effects.

For a homogeneous Poisson process with intensity λ, the intensity can be modelled as

$$\lambda = e^{\beta_0} \tag{1}$$

for any positive intensity such that β_0 is an intercept term representing the mean of the intensity. For an inhomogeneous Poisson process with spatially-varying intensity we can also add a Gaussian random field $\zeta(s)$ which accounts for spatial correlation in the observed points. This gives us

$$\lambda(s) = e^{\beta_0 + \zeta(s)} \tag{2}$$

where we can consider the exponentiated term a linear predictor such that $\eta = \beta_0 + \zeta(s)$. Again β_0 represents a mean intensity, but the random field $\zeta(s)$ captures the fluctuations about β_0. A log Gaussian Cox process is then a model where the point intensity λ can be described as:

$$\lambda(s) = e^{\eta(s)} \tag{3}$$

The observed spatially-varying intensity can also be modelled as

$$log(\lambda(s)) = \beta_0 + \sum_{m=1}^{M} \beta_m x_m(s) + \zeta(s) \tag{4}$$

where β_m are linear covariates that may influence the spatially-varying intensity of the points. These explain the observed spatial variation in the intensity of the process, with the spatially-varying Gaussian random field $\zeta(s)$ accounting for the fluctuations in intensity that the deterministic covariates cannot fully explain. More complicated, non-linear functions can also be included in the linear-predictor. In this way, $\zeta(s)$ describes variation of point intensity that is not accounted for by other model components and therefore highlights the spatial areas in which the model components are not sufficient to describe observed spatial patterns of intensity.

2.1.2 Integrated nested Laplace approximations (INLA)

To fit an LGCP model in a Bayesian manner, it is necessary to estimate the posterior distributions of the model parameters $\boldsymbol{\theta}$. Traditionally, Markov-Chain Monte-Carlo (MCMC) methods may have been used, however these are time-consuming to fit and prohibit the use of complex models. Where MCMC takes many samples from a posterior distribution, the integrated nested Laplace approximation (INLA) method (Rue et al., 2009) makes use of a series of approximations to estimate posterior distributions. Without the need for many iterations and the concerns of convergence associated with MCMC, INLA is a computationally efficient alternative for Bayesian analysis for models such as the LGCP where a latent Gaussian component is assumed. INLA is less generally applicable than MCMC because of the requirement for the latent Gaussian structure. Taylor and Diggle (2014) argue that an MCMC approach allows for a more flexible model and better analysis of joint posteriors than INLA. Teng et al. (2017) discuss different Bayesian approaches to the analysis of LGCP models, including different implementations of the INLA model, including the approach with stochastic partial differential equations that we use in this paper. In this paper we have chosen to use INLA due to the speed and ease of application facilitated by recent software developments, allowing rapid model configuration and comparisons.

The basic idea is described in Rue et al. (2009). Some observed data x_i can be described by a parameter vector $\boldsymbol{\theta}$, and each element of $\boldsymbol{\theta}$ can be described by some hyperparameters $\boldsymbol{\psi} = \{\psi_i ... \psi_k\}$ in a hierarchical Bayesian model. The INLA method aims to evaluate the marginal posteriors for each element of the parameter vector $\boldsymbol{\theta}$, which can be written as

$$p(\theta_i|x) = \int p(\theta_i, \psi|x)d\psi = \int p(\theta_i|\psi, x)p(\psi|x)d\psi \tag{5}$$

and for each element of the hyperparameter vector, $\boldsymbol{\psi}$, which can be written

$$p(\psi_k|x) = \int p(\psi|x)d\psi_{-k} \tag{6}$$

where $d\psi\_k$ is all other components of ψ except k.

It is necessary to calculate the joint posterior of the hyperparameters $p(\boldsymbol{\psi}|x)$ to calculate the posterior marginal distributions of the hyperparameters (equation 6), and also to calculate $p(\theta_i|\boldsymbol{\psi}, x)$ in order to solve equation 5 for the posteriors of each of the parameters θ_i. INLA does this by using Laplace approximations, nested because they are required for both $p(\boldsymbol{\psi}|x)$ and $p(\theta_i|\psi, x)$. INLA makes a Gaussian approximation of $p(\psi|x)$ which can be written $\tilde{p}(\psi|x)$, and a simplified Laplace approximation using a Taylor expansion of the approximation of $\tilde{p}(\theta_i|\psi, x)$. The main limitation of such an approach is the use of the Laplace approximation, which assumes that a smooth, peaked posterior distribution can be approximated by a Gaussian.

The posterior marginals can then be approximated with

$$p(\theta_i|x) \approx \int \tilde{p}(\theta_i|\psi, x)\tilde{p}(\psi|x)d\psi \tag{7}$$

which can be solved numerically through a finite weighted sum. This is a suitable approximation when the posterior marginals are roughly Gaussian in nature, but can also accommodate less-Gaussian posteriors, as discussed at length in Rue et al. (2009) and Rue et al. (2017).

2.1.3 Gaussian fields and Matérn correlation

Gaussian fields are a useful mathematical concept that can be used to model underlying or latent processes. In the LGCP framework outlined here, a Gaussian field ζ is used to model spatial variation not accounted for by the deterministic model components, β_m, as in equation 4. In this way, the Gaussian random field models the spatial structure by accounting for any spatial correlation between events. The combination of the random field and deterministic covariates models the intensity of the LGCP. We define the Gaussian field ζ as

$$\zeta(s) \sim GaussianField(0, \Sigma) \tag{8}$$

where the mean = 0 and the covariance is Σ. Calculating the covariance can be tricky, so instead of calculating all the values independently, a standard correlation function can be used to describe the correlation between points, and the area over which such correlation extends. A Matérn correlation function can be used to define the covariance such that

$$\Sigma = Cov_M = \sigma^2 Corr_M \tag{9}$$

where σ^2 is some variance parameter and Cov_M and $Corr_M$ are the Matérn covariance and correlation respectively. The Matérn correlation is specified as

$$Corr_M = \frac{2^{1-\nu}}{\Gamma(\nu)}(\kappa\|s_i - s_j\|)^\nu K_\nu(\kappa\|s_i - s_j\|) \tag{10}$$

where s_i and s_j are the spatial positions of observations i and j, and $\|s_i - s_j\|$ describes the Euclidean distance between these points. K_ν is a Bessel function, and the correlation has parameters κ and ν. Assuming that $\nu = 1$, the equation simplifies to have dependence only on κ and the distance between points. The Matérn correlation function describes the distances over which points within the model have some correlation, such that if the parameter κ is smaller there is more long-range spatial dependency.

With this approach, the parameters required to describe the underlying Gaussian random field are simply σ^2 and κ. This will still be time consuming to compute, unless we make the assumption that variation within the random field will only be on a local scale. If we can make the assumption that the underlying field is Markovian, such that only neighbouring points will have non-zero correlation, the correlation matrix becomes sparse. Such an assumption approximates the random field with a Gaussian Markov random field (GMRF). (Lindgren et al., 2011) provided an explicit link between Gaussian random fields and Gaussian Markov random fields that allows Gaussian random fields with Matérn covariance to be approximated by GMRFs even in cases where the spatial correlation structure is long range. The Matérn correlation structure is an extremely popular and flexible correlation structure used widely in a variety of spatial modelling applications (Guttorp & Gneiting, 2006).

2.1.4 Stochastic partial differential equations (SPDE) and mesh construction

The INLA approach can be used with continuous domain random field models as described by Lindgren et al. (2011) and Simpson et al. (2012), leading to the application of the method to a range of complex spatial and spatio-temporal models (Blangiardo & Cameletti, 2013; Gómez-Rubio et al., 2015; Lindgren & Rue, 2015). Essentially, this requires defining a mesh on which to construct the point process model, such that the random field can be evaluated at each mesh vertex. Lindgren et al. (2011) detailed how

random field models can be described by solutions of sets of stochastic partial differential equations (SPDEs). The parameters of the SPDEs are directly linked to the parameters of the Matérn correlation, so solving the SPDEs gives the required parameters for the Matérn covariance of the underlying GMRF. The SPDEs can be solved using a finite element approach, where the area can be represented by a mesh, with basis function representations used to calculate the value at each mesh vertex. Thus the SPDE approach allows the mapping of a random field from discrete points to a continuous field by the use of a correlation matrix that describes how the data points interact, or specifically the range of interaction of the points, which will be reflected in the resulting spatial intensity. This therefore allows us to consider a spatially continuous field rather than discrete point information, which in an earthquake context allows us to infer something about areas which have not experienced earthquakes within any point data set and not just the areas which have. A continuous field model also allows us to see where the model performs well in terms of the deterministic covariates describing the intensity, compared to areas where the intensity is mostly described by the random field component. Simpson et al. (2015) proved that the SPDE method converged well for log-Gaussian Cox processes, and with minimal error in the posterior distribution due to the method.

The mesh for the SPDE calculations is constructed using a restrained refined Delauney triangulation of a point dataset using the inla.mesh.2d function. The mesh boundaries are determined by the extent of the point locations, with a coarser mesh outside extending slightly outside of this area to reduce edge effects. A more complex mesh may provide greater resolution but will require more computational power, so a compromise is required which will provide reasonable resolution at an acceptable cost. The mesh can be constructed from the point locations themselves, but this results in a finer mesh in areas of clustering and a coarser mesh in areas with few events, when the spatial structure of most interest is likely to be somewhere in between these two extremes. As such we have chosen to specify the mesh domain as the spatial extent of the points rather than construct the mesh on the points themselves. This also makes models on different time periods more comparable where the point locations are likely to be different. A further consideration when constructing the mesh is the range of the Matérn correlation describing the random field. The correlation range in the remaining spatial structure must be greater than the length of the mesh edges so that the resulting intensities are reliable.

We construct and run the following models using the r package inlabru (Bachl et al., 2019) to fit LGCP models to the observed points using INLA. The inlabru package provides a user-friendly approach for using INLA for point process models, building on the R-INLA package (Rue et al., 2009; Lindgren et al., 2011; Martins et al., 2013). inlabru makes use of the sp package (Pebesma & Bivand, 2005; R. S. Bivand et al., 2013), using spatial data frames to handle data, and r packages raster (Hijmans, 2019), rgdal (R. Bivand et al., 2019) and rgeos (R. Bivand & Rundel, 2019) are also used for data wrangling. All maps in this paper are constructed with the use of ggmap (Kahle & Wickham, 2013) and tmap (Tennekes, 2018), with colour schemes from RColorBrewer (Neuwirth, 2014). The process for fitting LGCP models in inlabru is straightforward. A model is constructed for the random field component, based on a user-defined mesh. An equation describing the model components is defined, and an LGCP is fitted to this model. The LGCP fits can then be compared using DIC or by predicting the intensities that would be returned by the LGCP model. The predicted intensities can be a combination of all model components, or include only some of the modellers choosing. In this way, the effect of adding different model components can be compared by studying changes to the predicted intensity of the model as a whole and also by looking at the random field alone, which will identify spatial variability that the deterministic components cannot explain.

2.2 Model comparison and DIC

The deviance information criterion, or DIC, was developed by Spiegelhalter et al. (2002) as an alternative to the commonly used AIC. DIC is a measure that can be used to compare different models with varying numbers of parameters, designed as analogous to AIC for use with hierarchical models reflecting the trade-off between the 'goodness of fit' and model 'complexity'. The DIC is the expected deviance, penalised by the effective number of model parameters, which is a measure of the difference in posterior mean deviance and the deviance of the posterior means of the individual parameters. This penalises more complex models similarly to AIC, therefore preferring the simplest models that can explain the resulting data. DIC is designed specifically for hierarchical models where the model is structured in such a way that there is structural dependence between parameters, as is the case when we include parameter priors. The DIC is calculated within the INLA and inlabru software (and similarly within other software used for Bayesian hierarchical modelling such as BUGS), making it a popular choice for model comparison in these contexts (Spiegelhalter et al., 2014).

In this case the DIC is calculated at the posterior mean of the latent field and the posterior mode of the hyperparameters. The deviance D is defined by

$$D(\theta) = -2\sum_i \log p(x_i|\theta) \tag{11}$$

The effective number of parameters pD is calculated using the trace of the prior precision matrix Q multiplied by the posterior covariance matrix Q^*

$$pD = n - tr\{Q(\theta)Q^*(\theta)^{-1}\} \tag{12}$$

where n is the number of observations, and the total model DIC is then

$$DIC = \overline{D(\theta)} - pD \tag{13}$$

Rue et al. (2009) describe the full details of how DIC is calculated within the INLA software.

The Collaboratory for the study of earthquake predictability (CSEP, e.g. Michael & Werner, 2018) aims to improve earthquake forecasting by testing earthquake forecasts in real time, with the future aim of extending this approach to full hazard models (Schorlemmer et al., 2018). CSEP tests earthquake forecasts using a variety of different tests, ranging from simple tests of the number of forecast events (the N-test), to comparisons of forecast likelihoods (L-test) and residual comparisons, with different testing centres using different testing approaches for the models they are assessing. The DIC is also a likelihood-based model assessment tool, making it similar in some ways to the CSEP L-test, except in this case we are applying the tests to spatial data patterns rather than prospective or pseudo-prospective data at this time. We also compare the number of events expected by each of our constructed models - the abundance of points obtained by integrating over our mesh intensity models - which is somewhat analogous to the N-test used by CSEP. Nevertheless, the optimization presented here is a necessary but not sufficient criterion to developing a true prospective forecasting model. In future work we will develop pseudo-prospective and ultimately prospective forecast models tests to allow comparison with competing models in the CSEP framework.

Throughout this work we use DIC as a tool for model comparison as a first test, and highlight other methods of comparing models and model outputs. Our aim is not to discriminate between models based solely on their DIC, but the DIC and number of events predicted by the LGCP model make a useful first pass for testing and comparing models.

3 Data types

A huge benefit of the inlabru approach is in the ability to combine different data types within a model. The earthquakes themselves are described as points, and the deterministic components of the model can be included as lines, polygons, maps or raster images with discrete or continuous variables. Constructed LGCP models can include any combination of these components, and the output of one model can be straightforwardly included in the next. Continuous variables can be included with the use of a function that returns the value of the variable at a given point in space. Categorical information can be added for data which takes the form of discrete layers. In the inlabru terminology these are termed 'factor' covariates, and we demonstrate their use in constructing a binary 'fault factor' below. We begin by outlining the different data sets that can be included in an inlabru model with application to several datasets for California. Further information on each of the data types is included in the supporting information.

3.1 Earthquake Catalogue: Spatio-temporal point dataset

We make use of various sub-sections of the UCERF3 catalogue which consists of events above a fixed magnitude threshold. An LGCP model aims to model spatially varying intensity. The simplest possible LGCP model is one where we assume no known underlying spatial structure such that the intensity is a function of a Gaussian random field only. The smoothed seismicity model in Figure 1 is constructed in such a way, using a Matérn covariance for the random field. We see that the intensity model is behaving as we would expect, with high intensity in areas with a greater number of events.

3.2 Fault maps: Polygon dataset

Given that we know that the spatial distribution of observed earthquakes is related to underlying fault systems, we can include fault polygons in the model to see how well the fault locations account for the spatial distribution of events. The polygons are buffered as in the UCERF3 model. This is also the basis for our fault distance map which simply returns the distance to the nearest map fault for any point within the area of interest.

Alternatively, we consider two further covariates related to fault geometry. Since a significant number of events within the catalogue do not fall within the fault polygons, we can construct a fault-distance map which shows the distance from the nearest fault. As an alternative to the UCERF3 fault geometry, we also make use of the USGS Quaternary faults model (https://earthquake.usgs.gov/hazards/qfaults/background.php), which we shall refer to as 'QFaults'. Figure 1 shows the different inputs for the models in this paper, including the two different fault geometries, the fault distance model and the other spatial components described below.

3.3 Slip rate data: Spatial covariate or fault mark

There are four possible slip rate models considered within the UCERF3 logic tree. To work with these, we can construct either factor or continuous maps of slip rates for the different fault polygons, where a factor map requires discrete levels of data. In this paper, we use continuous slip rate values for each individual fault, such that the $\log_{10}$ slip rate is returned for any given point. Off-fault, the slip rate will be zero, so that we have essentially attached a value to each fault polygon only. Within the slip rate models, the value of the slip rate at any given point will be returned instead of the binary classification used to identify if a fault is present. The slip rates for the NeoKinema model used throughout this work are shown in Figure 1.

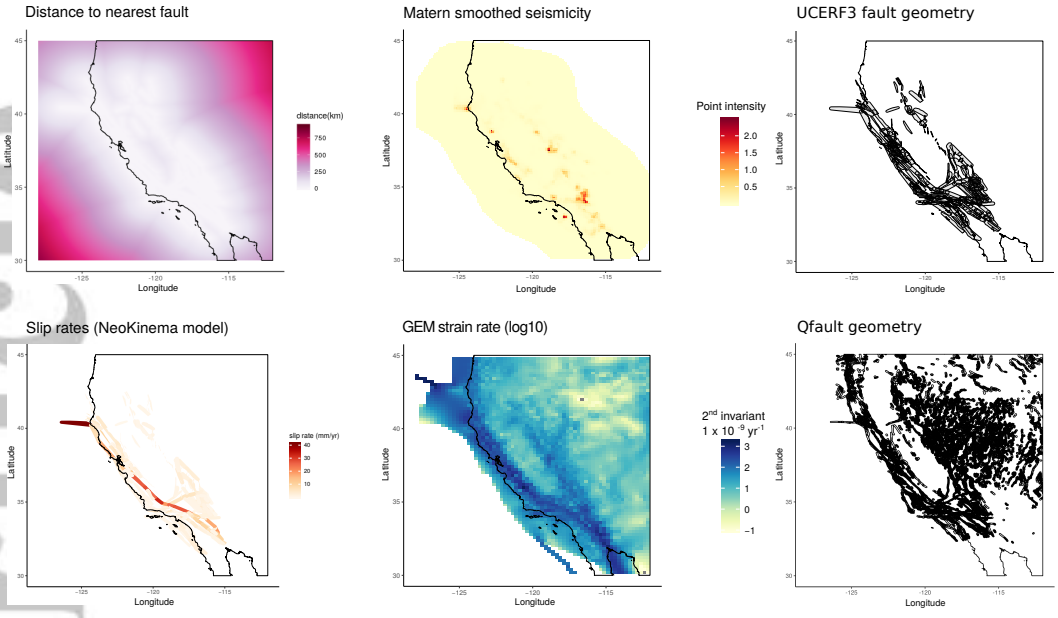

Figure 1. Input model components used within this work. Slip rates and fault geometry from UCERF3, Matérn smoothed seismicity created from UCERF3 $M2.5+$ dataset, distance to nearest fault calculated from UCERF3 geometry. Strain rates from GEM strain rate model, where we have used the $\log_{10}$ of these values. Qfault geometry from the USGS Quaternary fault model, cropped around the study area.

3.4 Past seismicity: Continuous spatial covariate

We use a subset of the UCERF3 data to construct a Matérn-smoothed past seismicity model, by fitting an LGCP model to the point data alone and predicting the model intensity for events that occurred before our time period of interest. This allows us to use the observed past seismicity as a spatial covariate in future models, where we have smoothed the past seismicity by assuming the intensity is a function of a random field only. We can therefore construct a past seismicity map using any subset of the data such that the past seismicity does not include any events in the model itself. A past seismicity model for all $M2.5+$ created using Matérn smoothing is shown in Figure 1.

3.5 Strain rate: Continuous spatial covariate

We make use of the GEM strain rate model (Kreemer et al., 2014) which is a global strain rate model constructed with the use of deforming cells in areas of high strain. Since the UCERF3 data is not global and instead from a small area with a good catalogue, the combined model of past seismicity and strain rate may perform less well than the past seismicity alone over the short timescales considered here, especially considering the resolution of the strain rate map. Over longer timescales the strain rate map may prove more useful by adding information. Given the abundance of model inputs we have access to in California, the strain rates may or may not be beneficial to the model, but the INLA method allows us to compare the effect of including a strain rate component with the addition of more detailed fault information. This allows us to assess if the strain rate might contribute meaningfully to models for areas which lack fault slip rates. The GEM strain rate model for California is shown in Figure 1.

Table 1. DIC estimates for models of synthetic data uniformly sampled from fault polygons.

Model	DIC	ΔDIC	Abundance
Random field + fault geometry	7178	0	498 ± 21
Fault geometry	7182	4	501 ± 23
Random field + slip rate	7785	607	510 ± 23
Random field only	7785	607	513 ± 23

4 Inversion of synthetic catalogues to demonstrate the method

To demonstrate the capabilities of the inlabru method, we construct two synthetic datasets based on the fault geometry of UCERF3. Using the R package sp (Pebesma & Bivand, 2005; R. S. Bivand et al., 2013), we randomly sample a chosen number of points from the fault polygons. The first model uniformly samples from within the fault polygons while the second weights the number of points from each fault polygon by the slip rates. We set the number of randomly sampled events from polygons to 500, but for the slip-rate weighted synthetic data the number of events varies and is much smaller. To properly assess the models, we construct 50 random datasets for each model. To each of these synthetic datasets we then fit five models: a random field model, a model with fault polygons only, a model with fault polygons and random field, a fault model with slip rates and a fault model with slip rates and random field.

4.1 Inversion of events randomly distributed on fault network

Events are randomly sampled from the buffered fault polygons using the spsample function. Figure 2 (left, top row) shows one random catalogue generated in this way and the resulting intensity predictions from the four constructed models. Table 1 shows the resulting model DIC values and predicted number of events from the LGCP fit as mean ± standard deviation for the dataset in Figure 2. The fault geometry model significantly improves upon the random field only model by accounting for much of the spatial distribution, but because the distribution within the fault polygons is random, the random field + fault geometry model performs the best. The model with slip rates performs better than the random field alone, as the slip rates are related to the fault locations, but as the slip rates do not describe the observed pattern of events in space, the DIC of the slip rate + random field model is higher than that of the fault geometry model. If we compare the number of events predicted by each of the four models, all of the models predict a reasonable number of events, given that 500 synthetic events have been used. The density plot in Figure 3 shows that the performance of the different models varies significantly with different randomly sampled events. The model with fault geometry overlaps almost entirely with the fault geometry and random field model, while the slip rate and random field and random field only models also overlap significantly. The slip rates alone do poorly in all fifty randomly sampled catalogues.

4.2 Inversion of events randomly distributed on fault network weighted by slip rate

For our second synthetic model, we construct a Log-Gaussian Cox model where the intensity consists of a Gaussian random field and the fault slip rates, where the slip rate component is similar in magnitude to the random field. This model is shown in Figure 2(f). The points are sampled using a built-in inlabru function (sample.lgcp), which samples around 160 events in each realisation, but the exact number varies. An example of the DIC results are shown in table 2, with these results corresponding to the intensities shown in Figure 2. The intensity range is extended by the inclusion of the slip rate model

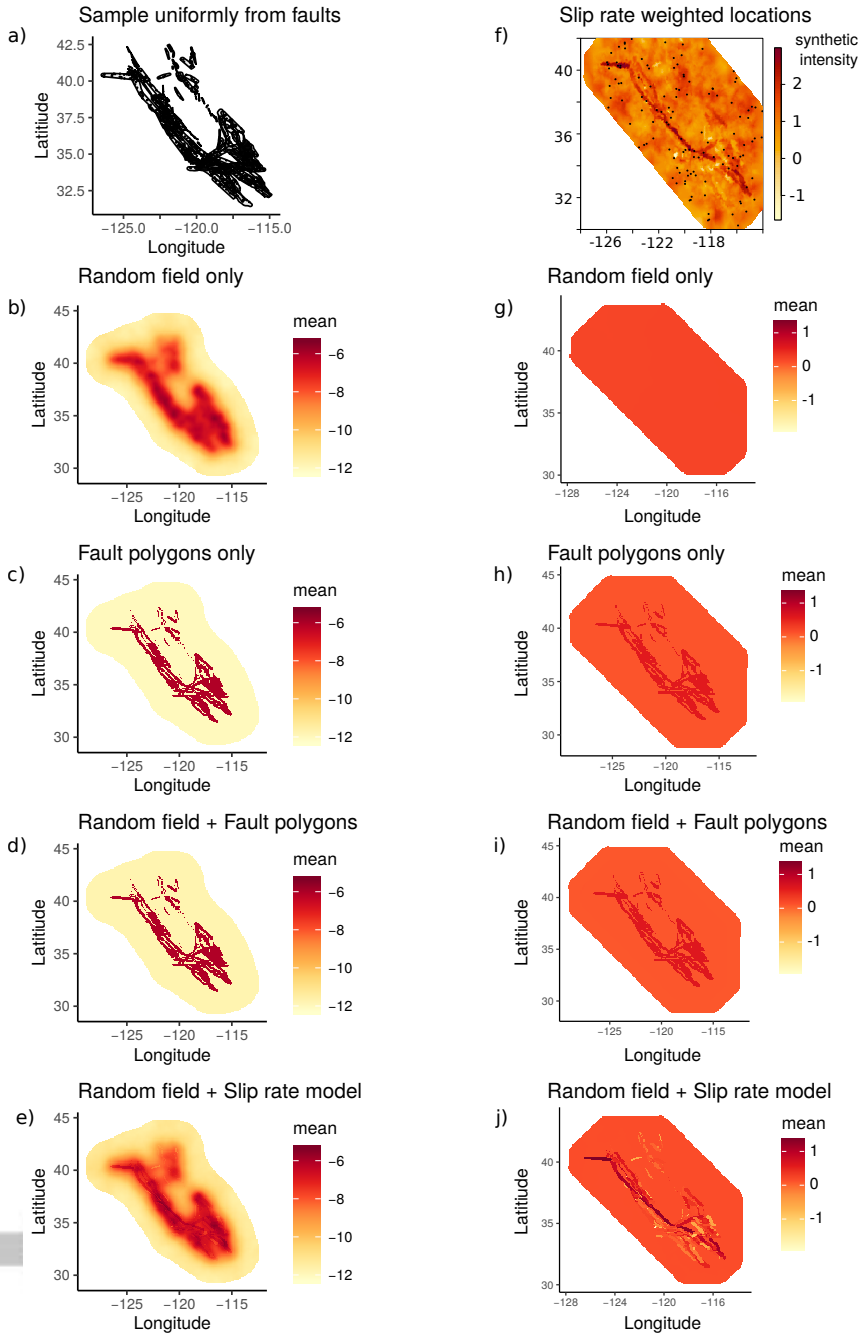

Figure 2. Inversion of two synthetic datasets: event locations randomly sampled from the full fault polygon set (a-e) and event locations sampled from a model of random field + fault slip rates according to the NeoKinema slip rate model (f-j). The top row shows the locations of synthetic events within the fault polygons. Plot (f) shows an example of the synthetic intensity model, where the fault polygons contribute to different intensities and the scale bar shows the log intensity of the synthetic model. Subsequent rows show predicted intensities for models of random field only (row 2, b & g); models which only include fault polygons (row 3, c & h); models with random field and fault polygons (row 4, d & i); and models with fault slip rates attached to the fault polygons and with included random field (bottom row, e & j). The scale bar shows the log posterior mean intensity.

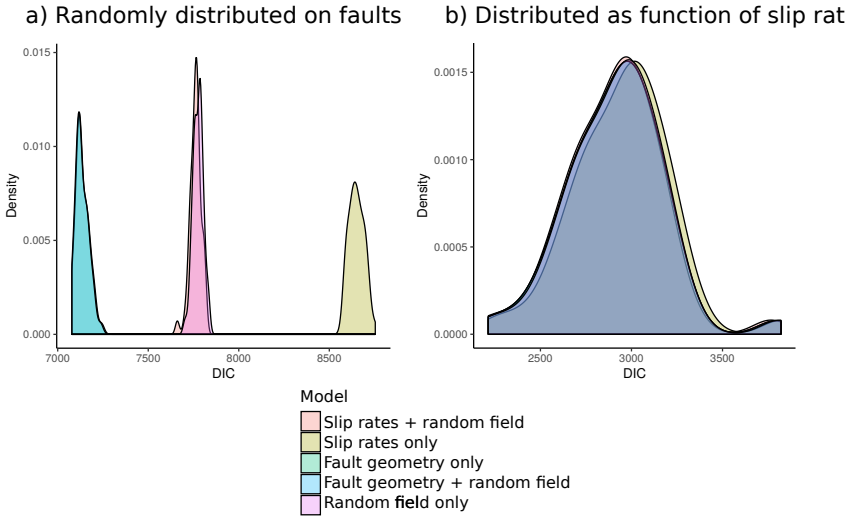

Figure 3. Densities of model DIC over 50 random samples from the fault polygons (a) and weighted by the slip rates (b)

Table 2. DIC estimates for models of synthetic data sampled as a function of slip rates.

Model	DIC	ΔDIC	Abundance
Random field + slip rate	3143	0	156 ± 14
Random field + fault geometry	3145	2	156 ± 10
Fault geometry only	3145	2	157 ± 12
Random field only	3155	12	155 ± 12
Slip rates only	3191	48	135 ± 13

(j), making the narrower ranges of models g-i appear almost uniform on the same scale. The number of modelled events in this sample was 155, with all but the slip rate only model giving a reasonable estimate of the number of events. It should not be surprising that the fault geometry model also works well, given that the slip rates are each associated with faults.

In this case our model is constructed based on the fault slip rates and a random component, so we would expect the fault slip rates + random field model to perform best. The densities in Figure 3 show how the resulting DICs overlap significantly in this model, with the slip rate + random field model outperforming the fault model + random field model very slightly. The fault geometry and slip rate models without random field also perform well on some occasions, because the slip rate component of the intensity is larger than the random field and the slip rates are only associated with faults. This results in all 5 models having similar DIC values, because all 5 of the models go some way to explaining the observed spatial distribution of events. In this case the small sample size may also contribute to the similar performance of each of the models, with the low number of samples making it difficult to identify which model performs best.

In both cases, the resulting DICs show a preference for the models including the underlying processes used to generate the dataset. In the randomly sampled model, the fault geometry model and fault geometry with random field are clearly the favoured models, with overlap resulting from the random generation of events. For the slip-rate weighted points there is more variation, arising from the way in which the synthetic datasets are

constructed allowing more variation in the resulting catalogues. It is clear, however, that the inlabru method is able to identify models which perform well and are consistent with the underlying patterns used to create the datasets in these synthetic examples. Such consistency is a necessary (but not sufficient) criterion in assessing forecasting power Murphy (1993). We therefore feel confident in moving forward and applying the method to real data where the true model is not known *a priori*.

5 Inversion of California dataset

5.1 UCERF Fault models

We first consider the fault geometries shown in Figure 1, using M5.5+ events. Figure 4 shows the resulting intensities for five different models for real data in California: one with only a random field (top), one with each of the fault geometries only and one for each of the fault geometries with an included random field. The fault polygons are included in the model as a binary factor covariate, such that the model checks if a fault is present at any point in space but does not consider any of the other fault information at this point. Because of this binary approach, all faults have an equal weighting within the model, hence why the middle row shows all faults as the same colour. Including a random field in the models along with the fault geometry (bottom) allows the model to account for events which do not occur on a fault polygon, or where there is a significantly high number of events that the fault polygons cannot explain.

The DICs for each of these models are shown in the top part of Table 3. The random field and fault geometry maps both have a lower DIC than the random field alone, suggesting that the inclusion of the fault maps improves the model's ability to account for the locations of the events. The fault geometries alone are not as good for describing the spatial location of events, while including the random field allows the models to account for extra spatial variation. We can also use the models to predict the number of events expected, given the fitted LGCP. These are shown as abundances in Table 3, where the mean and standard deviations are reported. There are 385 events in the UCERF3 M5.5+ catalogue, so the random field and fault geometry models predict very good numbers of events. The random field alone also predicts the correct number of events within one standard deviation, as do the fault geometry only models. By including a random field, we are accounting for the extra spatial variation that the fault map is missing, either as a consequence of the incompleteness of the fault map, the clustering of events or some combination of both.

The fault polygons clearly improve the intensity model when combined with a random field, but to what degree? To investigate this we applied models to each of the fault geometry buffers (see Figure S2 in supporting information), for both UCERF3 fault geometries, resulting in a total of 16 models, with half including a random field. The DICs for these models are reported in supporting material. Regardless of the buffer applied, the fault geometry and random field models perform better than the random field alone, with the fault geometry only models all performing worse. The combined buffer polygons perform better than the unbuffered polygons, or the uniform 1km buffer models, but the slightly better model appears to be the dip-dependent buffer, as this already does a good job of describing the spatial distribution of M5.5+ events. Fault geometry 3.2 performs better than 3.1 because it accounts for the spatial locations of the events slightly better - 240 of the 385 $M5.5+$ events occur within the fault polygons of FG3.2, compared to 237 for fault geometry 3.1. To simplify our model testing, we use fault geometry 3.2 for all further fault models, with the combined buffer applied.This allows us to use the UCERF3 slip rates, even though the QFault geometry is the better performer according to Table 3.

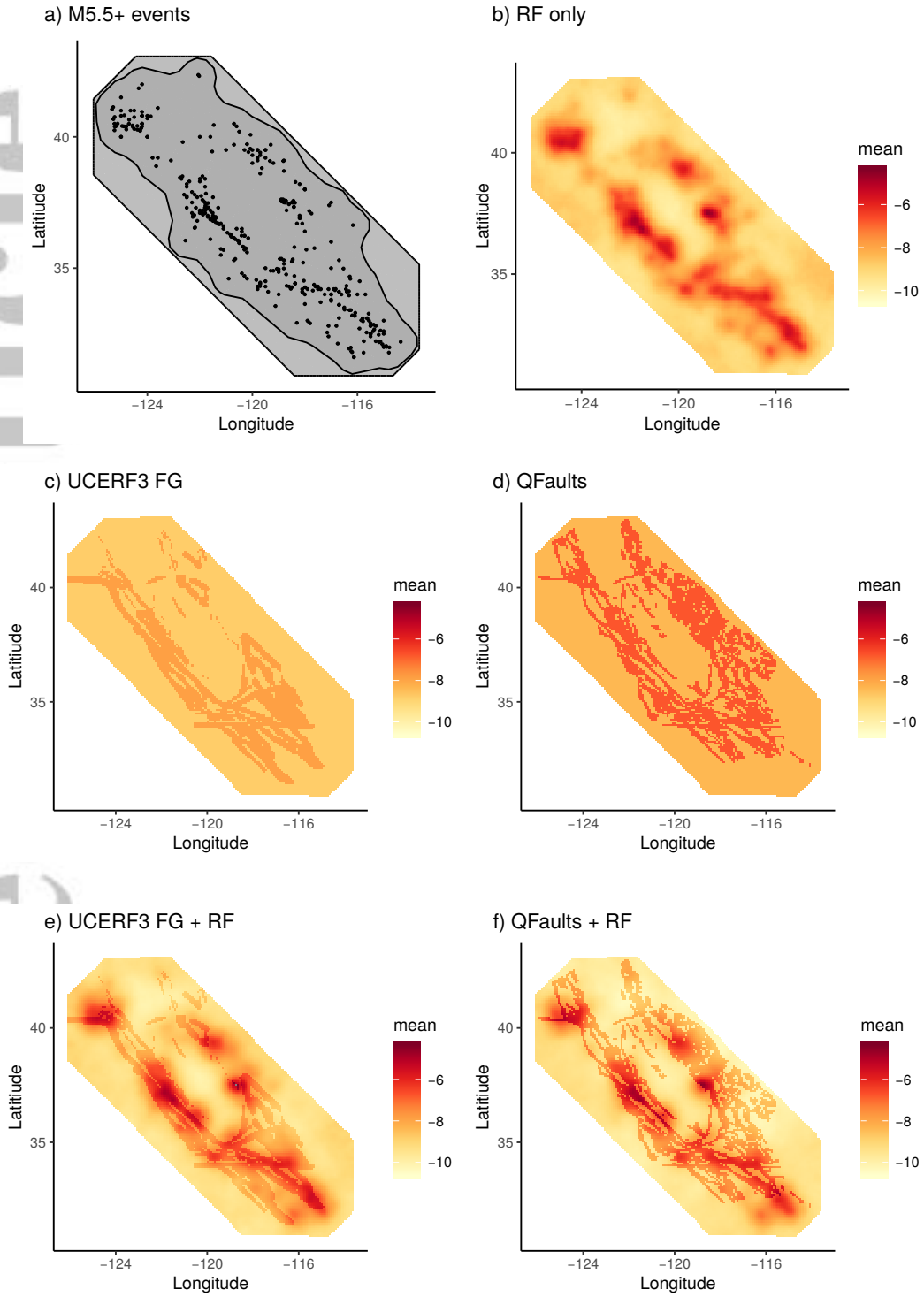

Figure 4. Inversion of fault models for two different fault geometries: UCERF3 fault geometry 3.2 and the Quaternary fault model (Qfaults). The top row shows the mesh used and locations of M5.5+ events (a) and the inversions for the model which includes only the random field (RF, b). The middle row (c, d) shows the resulting intensities for models of each of the two fault geometries (FG), where these have been included as a binary factor (events are on a fault or not). The bottom row (e,f) shows resulting model intensities for models which include fault polygons for each geometry and a random field. The scale bar shows the log posterior mean intensity.

We can also add slip rates for each fault according to one of four slip rate models in UCERF3. A comparison of the four slip rate models demonstrates that the NeoKinema slip rate model performs best of the four models in terms of DIC (see supporting material, Figure S3). Using the M5.5+ data, we see that the DIC for the slip rate + random field model is lower than that of the random field only or random field + fault geometry models, showing that the slip rates benefit the model.

5.2 Combining components

The true power of the inlabru approach is in the ability to construct models with different elements and compare their performance in terms of accounting for the spatial distribution of observed earthquakes. We construct 23 models containing combinations of the elements discussed above and shown in Figure 1, using the NeoKinema fault slip rates and fault geometry 3.2. The past seismicity model and distance to fault maps are created using events from 1984 to 2004 with $M \geq 2.5$, while all of the above models are for earthquakes occurring between 2004 and 2011 in the UCERF3 catalogue. The DIC for these models is much higher than when using the $M5.5+$ catalogue due to the greater number of involved events. The mesh used for each model is constructed based on the entire UCERF3 catalogue to provide adequate spatial coverage, though the mesh used for the past seismicity is extended slightly to avoid artefacts at the edges of the mesh in the later models. In the following discussion, past seismicity refers to a map of Matérn-smoothed past seismicity (see Figure 1) and the fault factor refers to a binary factor covariate which returns 1 in the event that any given point is within a fault polygon and 0 elsewhere, therefore representing a fault map.

Models with random field components perform significantly better than those without. The fault-distance is a more helpful inclusion for models with fault geometry than it is for models including slip rate. The fault-distance model provides similar information to the fault geometry model, but allows continuous variation with distance from the mapped fault location and so should be useful to both model types. This is potentially a consequence of the poor behaviour of the categorical fault factor rather than the inherent utility of the fault distance map per se. The fault distance map is also limited by the resolution that can be achieved within the map, currently around 2.5km. This resolution should be sufficient for the particular catalogue used here, where the majority of events outwith fault polygons have distances of ≥ 2.5km from the nearest fault polygon, but may be more challenging if many events are at very small fault distances.

The combined model with past seismicity + fault distance + slip rate + strain rate + random field performs best in terms of DIC. The model DICs tell us that past seismicity is more helpful than the fault distance when combined with slip rate + random field. Past seismicity is better able to account for areas of spatial clustering than fault distance maps, but the fault distance maps can account for some fault and event location uncertainty, so are helpful for the model in terms of improving the DIC. This suggests that the fault distance map adds extra useful information to the model, even when the fault map used to create it is included. To our knowledge this is the first time this result has been demonstrated.

The $M2.5+$ model contains significant clusters of events. To test if the model component performance is different at other magnitude cutoffs, we consider a catalogue of $M4.5+$ events, again using the split catalogue. The $M4.5+$ catalogue contains far fewer events, with only 127 earthquakes. Many of these are related to the $M7.2$ 2010 El Mayor-Cucupah earthquake in Baja California. The location of these events and the five best models by DIC for each of these two datasets are shown in Figure 5. The two different data sets have different scales because of the different number of events, though the resulting intensity patterns are similar. The past seismicity + strain rate model performs well for both datasets in terms of its ranking. The slip rates (NK) perform better for the

Table 3. DIC results for combined models with $M5.5+$ (top) and $M2.5+$ (bottom), where Δ DIC compares DIC for the 'best' model and δDIC compares DIC for the next best model.

M	Model	DIC	ΔDIC	δDIC
5.5	Random field + Qfault geometry	5849	0.0	0.0
	Random field only	5919	70	70
	Random field + UCERF3 fault geometry	5920	71	1
	QFault geometry	6296	447	376
	UCERF3 Fault geometry	6322	473	26
2.5	Strain rate + past seismicity + fault distance + slip rate + random field	54961	0.0	0
	Fault geometry + past seismicity + slip rate + random field	55176	215	215
	Past seismicity + strain rate + random field	55221	260	45
	Past seismicity + strain rate + slip rate + random field	55239	278	18
	Fault geometry + fault distance + past seismicity + random field	55255	294	16
	Fault geometry + past seismicity + random field	55347	386	92
	Fault distance + past seismicity + random field	55673	712	326
	Fault distance + past seismicity + slip rate + random field	55685	724	12
	Fault geometry + strain rate + slip rate + random field	55727	766	54
	Fault geometry + strain rate + random field	55758	797	31
	Past seismicity + random field	55979	1018	221
	Past seismicity + slip rate + random field	55992	1031	13
	Fault geometry + fault distance + random field	56185	1224	193
	Strain rate + slip rate + random field	56231	1270	46
	Strain rate + random field	56236	1275	5
	Fault geometry + random field	56373	1412	137
	Fault distance + random field	56472	1511	236
	Fault distance + slip rate + random field	56479	1518	7
	Slip rate + random field	56812	1851	333
	Random field only	56821	1860	9
	Fault distance only	109791	54830	52970
	Past seismicity only	112361	57400	2570
	Slip rate only	112697	57736	336

$M4.5+$ dataset due to the higher slip rates on faults in the Baja California region, where the majority of $M4.5+$ events occur. This leads to the fault geometry being a stronger constraint on the resulting intensity as opposed to in the $M2.5+$ catalogue.

As well as looking at the performance of different models spatially and their DIC, we can examine the posteriors of the different model components to see how much each component contributes to the intensities. This allows us to consider the contribution of individual components to the models and how this changes with different component combinations. This is especially useful when we are using components that may account for similar spatial patterns.

In the case of the $M2.5+$ model with the lowest DIC, the past seismicity and strain rate components contribute most significantly to the intensity with a posterior mean and standard deviation of 5.6 ± 0.2 and 3.76 ± 0.15 respectively, while the fault distance and slip rate components have a much smaller contribution at -0.076 ± 0.005 and 0.009 ± 0.007. The full posterior distributions can be found in supporting Figure S4. Table 3 shows that though the contributions of the fault distance and slip rate are small, they do improve the DIC of the model overall.

In both datasets, the smoothed seismicity always improves the model DIC by accounting for some of the significant spatial clustering of events and by highlighting areas where many events have occurred before. As the past seismicity includes smaller events, it can therefore account for very high intensity in areas which have experienced large earthquake sequences in the past. The strain rate model also consistently performs well, both on its own and in combination with other components. The model of strain rate and past seismicity performs very well (Table 3), which is consistent with the Strader et al. (2018) assessment of the GEAR1 forecast. In this case, we can see from the output model intensity that the strain rate model is important for constraining the spatial limits of seismicity in a way that other model components cannot. This suggests that the strain rate can contribute helpful information to forecasts, and the global availability of the strain rate map means this can be added to models for other regions where detailed information on faults may be unavailable or have higher uncertainty. Though the fault geometry and fault distance maps are not particularly useful on their own, they improve models when they are included by accounting for some of the smaller spatial structure that is not defined by better performing components. The fault distance map may account for some of the uncertainty in the fault geometry, but it also improves models which otherwise only include the fault geometry, because it can account for the lack of events in the Northeast and Southwest of the map. The fault distance map is likely to be more useful in areas where the fault map is less complete or highly uncertain. It may also perform better when fault buffers are smaller or not constrained by fault dips. The slip rates proved to have variable performance, proving useful for $M \geq 5.5$ models (see also Supporting information figure S1) and in $M4.5$ models, but less useful for smaller events (Table 3).

Component performances are ranked according to their individual performance and their performance in combination with other components considering both data sets. The availability of the model components is also considered, as are any alternatives that could be considered in model construction if the specific component is unavailable or as a comparison. While the past seismicity is catalogue-dependent, it can be calculated for any region by fitting a random field to the point data. The fault geometry and slip rates are highly dependent on the area of study, but the fault distance map can potentially help to explain some of the spatial distribution of events where these components are incomplete or unavailable. The random field from models with fault geometry and fault distance components could help to identify areas in which the fault map is incomplete. The GEM strain rate model is a global strain rate model and is therefore available for any region, however the deforming cells at oceanic plate boundaries are poorly constrained which may lead to some model artefacts. This can be seen in our California example (Fig-

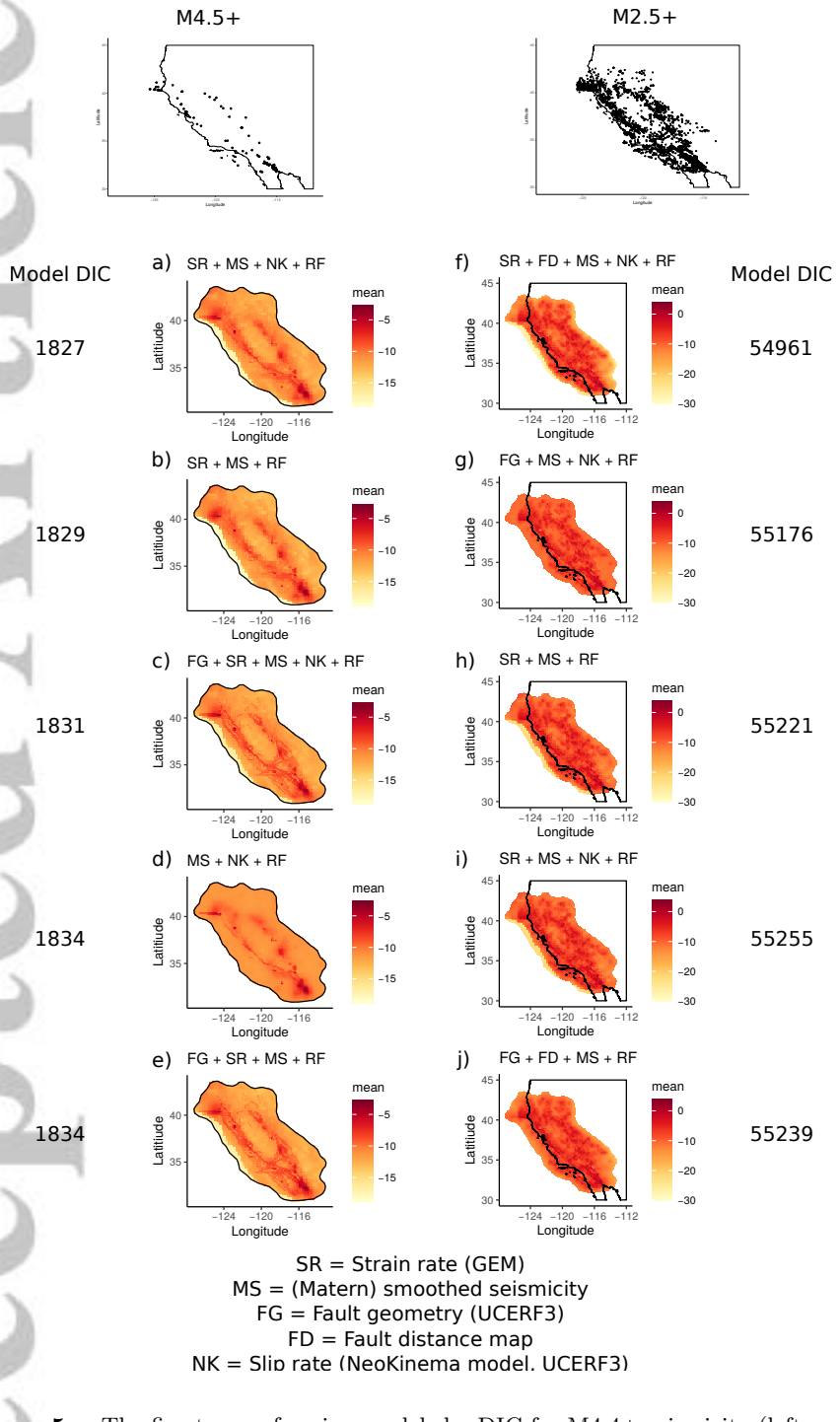

Figure 5. The five top performing models by DIC for M4.4+ seismicity (left, a-e) and M2.5+ seismicity (right, f-j). The scale bar shows the log posterior mean intensity. Intensity scales are different for the two datasets given the different number of events. Model component performance also varies depending on the dataset, with the fault slip rates proving more useful in the M4.5+ seismicity leading to the prevalence of fault structures in the intensity maps.

ure 1) as the thin blue stripe in the South of the map, which is the result of the plate boundary imposed by the modelling.

This provides us with a foundation for building models for any region, by identifying which components work well and which components can be used in their place should they be unavailable, which is summarised in Figure S5 of supporting information. Extending this approach by considering other data types and performance in other regions could prove valuable for constructing future earthquake forecasts in future work.

5.3 Fault geometry, slip rates and their effects at different magnitudes

We have considered so far the UCERF3 fault geometry and slip rate data, but the above comparisons of model component combinations at different magnitudes demonstrate that the slip rates may have different importance for different datasets. We test this by comparing slip rate and fault geometry models at different magnitude cutoffs. We also use a set of randomised slip rates, where the slip rates are randomly re-assigned to faults within the geometry, to test how much the usefulness of the slip rates is influenced by the actual slip rate values and how much is a function of the models ability to consider different faults more or less important. Finally we also include the QFault model as an alternative fault geometry. This combination of models allows us to assess the performance of slip rates and fault geometry as a function of the magnitude cutoffs and to further explore why different model components perform well or otherwise.

Table 4 ranks each of the five models at several different magnitude thresholds, highlighting changes to model ranking as a result of the changes in dataset. Lower magnitude thresholds will result in catalogues with more significant clustering, but as mentioned with the M4.5+ catalogue above, there may still be significant clustering even at higher magnitude cut-offs. Longer catalogues could help to minimise the effect of recent large sequences. The randomised slip rates will perform better when they better account for the observed seismicity at different cut-offs. In this case the randomised slip rates do not perform better than modelled slip rates at magnitudes $M4+$. This is consistent with the results for $M2.5+$ seismicity above, and suggests that the slip rates are more useful for describing the locations of large events. For events $M4.5+$, the slip rates perform better than the fault geometry alone, but in each of these models the Quaternary fault model performs better than the UCERF3 fault geometry. Figure 1 demonstrates that the Quaternary fault geometry includes many faults to the Northeast which can better account for seismicity in this region. The UCERF3 fault geometry performs better than the Quaternary faults at lower magnitudes due to the changing number and distribution of events. For larger magnitude catalogues, events not related to the UCERF geometry faults will be significant as a function of catalogue size, but this will be less true for smaller magnitude events (and therefore larger catalogues). This may also be related to the buffering of the fault geometries being different, with the fixed buffering of the Quaternary faults performing more poorly than the dip-dependent buffering in the UCERF geometry. The changes in model ranking at different magnitude cutoffs highlights the effect of clustering and the inclusion of smaller events on model component performance and the complexity involved in constructing such models.

In summary, models including the fault maps and strain rate data are always improvements on the random field seismicity model alone, and add more information for larger magnitude thresholds. This is important for applications to seismic hazard, which is normally dominated by intermediate to large magnitude events, i.e. if the frequency-moment distribution takes the form of a pure power law or a power law with an exponential taper (e.g. Main, 1995). In this case there may be a minimum threshold for being likely to be felt at a particular intensity likely to cause damage, often set at magnitude 5 for design of Nuclear Power Plants (Bommer & Crowley, 2017). In such cases,

Table 4. Fault model rankings at different magnitude cutoffs

Model	M5.5+	M5.0 +	M4.5+	M4.0+	M3.5+
UCERF faults	3	3	2	1	1
QFaults	1	1	1	2	3
NK slip rates	2	2	3	3	4
Randomised slip rates	5	5	4	4	2
Random field only	4	4	5	5	5

the analysis presented in Table 4 shows the fault map and strain rate data provide critical constraints in describing the seismicity.

6 Discussion

6.1 Smoothed seismicity

On short forecasting timescales, we should expect local clustering of earthquakes to dominate so we would expect recent smoothed seismicity models to be informative. On longer timescales we would expect the entire seismogenic region to be sampled so a longer sample may better account for longer-term trends in seismicity that are better captured by fault and strain rate maps. For time-independent forecasting, as long a sample as possible should be used, so that the effect of short-term clustering is reduced and as many large events as possible are captured within the past seismicity model.

The earliest models submitted to the Regional Likelihood Earthquake Model testing (RELM, a precursor to CSEP), were mostly constructed on the basis of some form of smoothed seismicity (Field, 2007), though some also included strain rates or geological information. The preliminary results for California found that the smoothed past seismicity model of Helmstetter et al. (2007) provided the best result of all submitted models (Schorlemmer et al., 2007; Zechar et al., 2013) whether aftershocks were included or not. Smoothed seismicity has also been found to perform well for forecasting when combined with strain rate (Bird et al., 2015; Strader et al., 2018) and when updated regularly as part of one-day forecasts in New Zealand (Rhoades et al., 2018). Our results above demonstrated the utility of the past seismicity in spatial seismicity models, both individually and in combination with other components. We saw that including the past seismicity with other components always improved model DIC compared to models with the same components and no past seismicity. We are therefore quite confident that our results using the inlabru approach are consistent with findings well-documented elsewhere in the literature. Further, we have demonstrated that a Matérn smoothing of the spatial intensity is also suitable for describing past seismicity (see supporting material and Figure S7 for explicit comparison). The improved performance of the Matérn smoothed seismicity in the inla model may be a result of the alternative gridding - the Matérn smoothing is based upon a Delauney triangulated mesh constructed from earthquake locations, rather than a uniform grid. We propose that this may well influence the effect of the smoothed seismicity within the model. In particular, this may be interesting in settings where gridded smoothed seismicity has proved less useful to forecast models, such as in the Italian CSEP tests (Taroni et al., 2018) which found recent smoothed seismicity to perform less well than a model which included more historic seismicity and fault locations. The inla framework would also allow the ranking of models containing each of these components as combining the components in one model is straightforward.

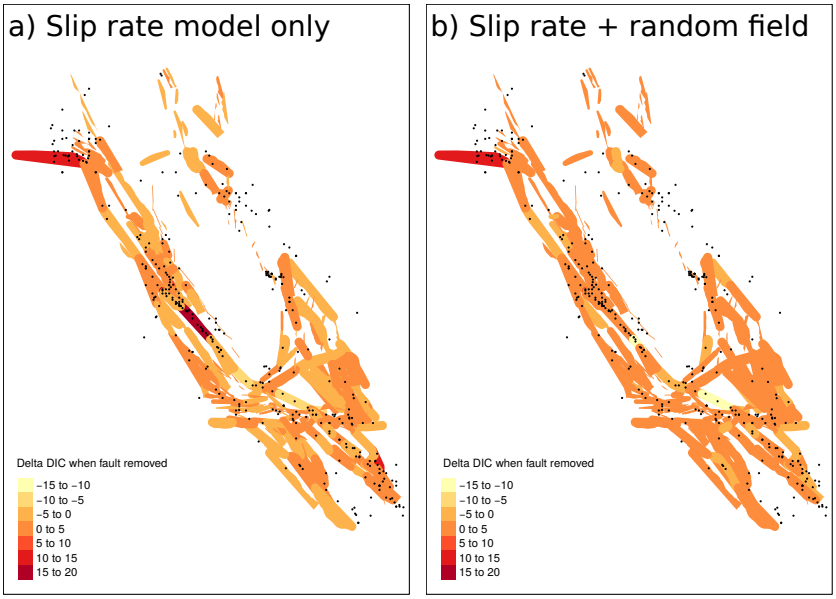

Figure 6. ΔDIC values for each fault in the UCERF3 catalogue, where the colour of the fault reflects the change in total model DIC when it is removed from the fault polygons. The locations of the $M5.5+$ events are shown in black, with the left panel (a) showing the results for a model of slip rate fault polygons only and the model on the right (b) including a random field component.

6.2 Slip rates in California

Our results demonstrated that the slip rate performance is quite variable but generally better for large events which are more likely to be independent. This would suggest that slip rates could be a valuable constraint for stationary, time-independent hazard models. To explore the full effect of slip rates on the model DIC, we remove each fault polygon from the model sequentially and record the DIC for each model to calculate a ΔDIC for each fault. We do this for two different model types to see the effect of including a random field and consider only events with $M \geq 5.5$.

Figure 6 shows the results for a fault model only (left) and for a model that includes a random field (right). High positive ΔDIC values suggest that a fault is important to the model, as removing it causes the total model DIC to increase relative to the model with all faults. Conversely, faults with negative ΔDIC values suggest that the overall model is improved by their absence. The models with fault slip rates alone return positive ΔDIC values for 102 faults, while the models with random field return positive ΔDIC values for 286 of 320 faults. This suggests that the random field model finds all faults more valuable to the model than a model that includes slip rate alone, such that removing any particular individual fault will have a similar effect on DIC, with a few exceptions. Removing even the largest faults has little effect on DIC, while removing some of the smaller faults is more significant due to the greater number of events associated with them, as shown in Figure S8.

The models that include the random field are more likely to have an increased DIC when faults are removed than the slip rate only model (Figure S8). Removing the faults from the model in this way does not affect DIC for the models for 95 models without random field component, but none of the faults have $\Delta\text{DIC} = 0$ when a random field is included in the model, suggesting all faults have a contribution to the total DIC in the

Table 5. Seismicity models for July 2019 Ridgecrest sequence, where Δ DIC compares DIC for the 'best' model and δDIC compares DIC for the next best model

Model	DIC	ΔDIC	δDIC
Past seismicity + random field	1006	0	0
Strain + random field	1097	91	91
UCERF faults + random field	1121	115	24
Quaternary faults + random field	1129	123	8

model with random field. Further, the faults with large ΔDIC for the models including a random field have a smaller ΔDIC than in models without a random field, such that removing any one individual fault will have a smaller effect on total DIC for a model with random field. It is also worth noting that the total DIC is always lower for models with a random field than models without. The random field is able to explain more of the spatial distribution than the slip rate model alone, such that removing any individual fault does not have a significant impact on resulting DIC. This is promising for areas where the fault map is less complete. These findings demonstrate that the slip rates in California cannot fully account for the observed spatial distribution of earthquakes. The slip rate or size of an individual fault does not correlate with the number of events occurring within the fault polygon.

6.3 The 2019 Ridgecrest sequence

The July 2019 Ridgrecrest sequence in southern California allowed us to apply the inlabru method to a recent event sequence. We used a catalogue of 1459 M2.5+ events that occurred between the 15th of June and the 15th of July 2019 retrieved from the USGS earthquake database. We consider events that fall within an area of -117 - -118 longitude and 35 - 37 latitude. Within our Ridgecrest catalogue, 985 events are not within a buffered UCERF3 fault polygon, with the remaining 474 events linked to 5 different fault polygons. The M7.1 event on the 6th of July and the M6.4 event two days previously are not within any UCERF3 buffered fault polygon, with the largest event directly linked to a fault polygon being a magnitude 5.5 event on the Airport Lake fault. This motivated us to try the USGS Quaternary fault model as an alternative fault map.

We constructed four models for seismicity during this time period, each including a random field and one spatial covariate: UCERF3 and Quaternary fault geometries,the GEM strain rate and a past seismicity model based on the entire UCERF3 catalogue and subsequent events (i.e. events from May 2012 - January 2019 as well as the UCERF3 events). All four models were compared to a random field only model and proved to improve the model DIC, highlighting that each of these covariates was helpful in describing the spatial patterns of seismicity.

The past seismicity and strain rate models perform best according to the DICs recorded in Table 5, though the predicted intensity models show the grid pattern that comes from the spatial resolution of the input covariates. This gridding appears to benefit the models, with high intensity areas falling within grid-squares with higher past seismicity. The area of the Ridgecrest events is a generally higher strain rate area than most of Southern California. In both cases, increased resolution of the data may improve the performance, but we can see that the current resolution is enough to improve the model. The UCERF3 geometry outperforms that of the Quaternary faults, but this appears to be a result of the buffering applied: the larger buffers of the UCERF3 model allow it to account for some of the events despite the lack of matching geometry (see supporting information figure S6). The Quaternary fault model contains many smaller faults and it

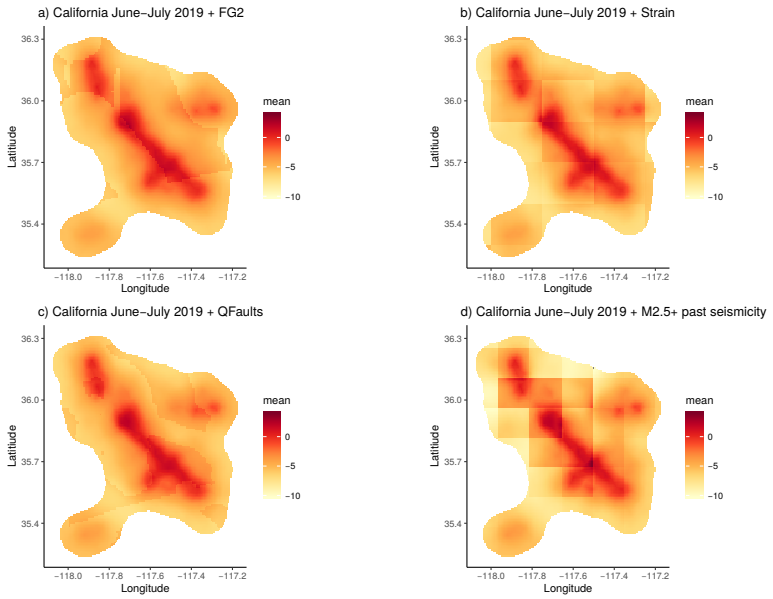

Figure 7. Predicted field intensities for the Ridgecrest events for four different models: a)UCERF faults, b) strain rate, c) Quaternary fault geometry and d) past seismicity. The scale bar shows the log posterior mean intensity.

is likely that a larger spatial buffer on these faults would fill the area of the Ridgecrest events without necessarily accounting for the correct geometry. In this case neither fault map performs as well as the strain rate, which reinforces the usefulness of the strain rate in spatial models of seismicity. Choosing appropriate buffers for fault projections will remain an on-going challenge for fault-dependent models, but by including other components in a model the effect of incomplete fault maps or poorly chosen buffers can potentially be reduced.

6.4 Current Limitations

This paper is the first attempt at applying inlabru for modelling seismic hazard based primarily on existing functionality. There are currently several limitations that we are addressing in ongoing work. This will include extending the model to include magnitudes of events as marks jointly modelled with the random field. Event magnitudes are likely to affect the spatial correlation range of events, thus modifying the structure of our random field. They have been considered here in terms of the choice of catalogue only, with the magnitudes of individual events not considered in the model. The model will then be further extended to include time-dependence, which requires the model to be self-exciting, for example applying an ETAS model to the data. Further extensions to account for uncertainty in event location and in the various input parameters will also be considered, as well as the development of a robust model assessment and comparison approach so as not to rely so heavily on the DIC and number of events alone. (Marzocchi et al., 2012) suggest that when constructing hybrid models the correlation between included forecasts should be considered when weighting model components. The components of our models are sometimes highly correlated, but the contribution from each component to the total mean intensity can be considered using the component posterior means. This allows us to identify under which circumstances the contributions of individual components are most significant to the observed intensity. Future work will make use of these component contributions in the construction of prospective forecasts of seismicity, and as-

sess the extent to which the inclusion of different (perhaps correlated) parameters affects the spatial distribution of forecast events. Nevertheless, we have presented a proof-of-concept for the inlabru method, demonstrating that it is a promising method that can be developed in future research. We have also used it to demonstrate several findings consistent with previous work, and some that are specific to the present work.

7 Conclusions

We have demonstrated for the first time that Log-Gaussian Cox process models for earthquake data can be constructed, fitted, and tested efficiently using integrated nested Laplace approximations in a purely data-driven approach. The inlabru approach and model framework allows the quick and easy construction of seismicity models that include various different types of information, including fault polygons (with or without some attached mark of their own), derived products such as distance to the nearest fault, and spatially-continuous models, such as strain rate data and past seismicity. The inlabru approach confirms results from the literature in terms of the inclusion of past seismicity and fault information in seismicity models, allows the straightforward combination of different data types and allows the ranking of models by making use of the model DIC. Including strain rate data constrains the spatial limits of seismicity in California, while maps of distance to nearest fault prove beneficial even to models which also include the fault maps themselves by accounting for seismicity not associated with specific mapped faults. Further, the importance of information within individual components can also be tested, such as by considering changes to the model DIC when individual faults are removed. Such a framework allows the user to identify which model components truly benefit the model, the best combination of different model components, and to identify spatial areas in which their model is currently lacking by considering the random field.

Acronyms

LGCP Log-Gaussian Cox process
INLA Integrated nested Laplace approximations
GMRF Gaussian-Markov Random Field
SPDE Stochastic partial differential equations
DIC Deviance Information Criterion

Acknowledgments

UCERF3 data is available online from https://pubs.usgs.gov/of/2013/1165/. KB was funded during this work by an EPSRC PhD studentship (grant number 1519006), and during the writing of this paper by Real-time Earthquake Risk Reduction for a Resilient Europe 'RISE' project, which has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 821115. We thank Francesco Serafini for feedback on the synthetic models and an early version of this paper. The authors thank Shyam Nandan and an anonymous reviewer for their thoughtful and constructive comments.

References

Bach, C., & Hainzl, S. (2012, 4). Improving empirical aftershock modeling based on additional source information. *Journal of Geophysical Research: Solid Earth*, *117*(4), n/a-n/a. Retrieved from `http://doi.wiley.com/10.1029/2011JB008901` doi: 10.1029/2011JB008901

Bachl, F. E., Lindgren, F., Borchers, D. L., & Illian, J. B. (2019, 6). in-

labru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, *10*(6), 760–766. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13168 doi: 10.1111/2041-210X.13168

Bird, P. (2009, 11). Long-term fault slip rates, distributed deformation rates, and forecast of seismicity in the western United States from joint fitting of community geologic, geodetic, and stress direction data sets. *Journal of Geophysical Research: Solid Earth*, *114*(B11). Retrieved from http://doi.wiley.com/10.1029/2009JB006317 doi: 10.1029/2009JB006317

Bird, P., Jackson, D. D., Kagan, Y. Y., Kreemer, C., & Stein, R. S. (2015, 10). GEAR1: A Global Earthquake Activity Rate model constructed from geodetic strain rates and smoothed seismicity. *Bulletin of the Seismological Society of America*, *105*(5), 2538–2554. doi: 10.1785/0120150058

Bivand, R., Keitt, T., & Rowlingson, B. (2019). *rgdal: Bindings for the 'Geospatial' Data Abstraction Library.* Retrieved from https://CRAN.R-project.org/package=rgdal

Bivand, R., & Rundel, C. (2019). *rgeos: Interface to Geometry Engine - Open Source ('GEOS').* Retrieved from https://CRAN.R-project.org/package=rgeos

Bivand, R. S., Pebesma, E. J., & Gomez-Rubio, V. (2013). *Applied spatial data analysis with R, Second edition.* Springer, NY. Retrieved from http://www.asdar-book.org/

Blangiardo, M., & Cameletti, M. (2013, 3). Spatial and Spatial-temporal Bayesian Models with R-INLA. *Spatial and Spatiotemporal Epidemiology*, *4*, 33–49. doi: 10.1016/j.sste.2012.12.001

Bommer, J. J., & Crowley, H. (2017, 7). The purpose and definition of the minimum magnitude limit in PSHA calculations. *Seismological Research Letters*, *88*(4), 1097–1106. doi: 10.1785/0220170015

Cox, D. R. (1955). Some Statistical Methods Connected with Series of Events. *Journal of the Royal Statistical Society. Series B (Methodological)*, *17*, 129–164. Retrieved from https://www.jstor.org/stable/2983950 doi: 10.2307/2983950

Díaz-Avalos, C., Juan, P., & Serra-Saurina, L. (2016, 12). Modeling fire size of wildfires in Castellon (Spain), using spatiotemporal marked point processes. *Forest Ecology and Management*, *381*, 360–369. Retrieved from https://www.sciencedirect.com/science/article/pii/S0378112716305746 doi: 10.1016/j.foreco.2016.09.013

Diggle, P. J., Moraga, P., Rowlingson, B., & Taylor, B. M. (2013, 11). Spatial and Spatio-Temporal Log-Gaussian Cox Processes: Extending the Geostatistical Paradigm. *Statistical Science*, *28*(4), 542–563. Retrieved from http://projecteuclid.org/euclid.ss/1386078878 doi: 10.1214/13-STS441

Dutra Silva, L., Brito de Azevedo, E., Bento Elias, R., & Silva, L. (2017, 12). Species Distribution Modeling: Comparison of Fixed and Mixed Effects Models Using INLA. *ISPRS International Journal of Geo-Information*, *6*(12), 391. Retrieved from http://www.mdpi.com/2220-9964/6/12/391 doi: 10.3390/ijgi6120391

Field, E. H. (2007, 1). Overview of the Working Group for the Development of Regional Earthquake Likelihood Models (RELM). *Seismological Research Letters*, *78*(1), 7–16. Retrieved from https://pubs.geoscienceworld.org/srl/article/78/1/7-16/143319http://srl.geoscienceworld.org/cgi/doi/10.1785/gssrl.78.1.7 doi: 10.1785/gssrl.78.1.7

Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., ... Zeng, Y. (2014). Uniform California Earthquake Rupture Forecast, version 3 (UCERF3) -The time-independent model. *Bulletin of the Seismological Society of America*, *104*(3), 1122–1180. Retrieved from

https://pubs.geoscienceworld.org/ssa/bssa/article-pdf/104/3/1122/2671948/1122.pdf doi: 10.1785/0120130164

Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., ... Zeng, Y. (2015). Long-term time-dependent probabilities for the third uniform California earthquake rupture forecast (UCERF3). *Bulletin of the Seismological Society of America*, *105*(2), 511–543. doi: 10.1785/0120140093

Field, E. H., Milner, K. R., Hardebeck, J. L., Page, M. T., van der Elst, N., Jordan, T. H., ... Werner, M. J. (2017, 6). A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bulletin of the Seismological Society of America*, *107*(3), 1049–1081. Retrieved from https://pubs.geoscienceworld.org/bssa/article/107/3/1049-1081/354189 doi: 10.1785/0120160173

Gómez-Rubio, V., Cameletti, M., & Finazzi, F. (2015, 11). Analysis of massive marked point patterns with stochastic partial differential equations. *Spatial Statistics*, *14*, 179–196. Retrieved from https://www.sciencedirect.com/science/article/pii/S2211675315000512 doi: 10.1016/j.spasta.2015.06.003

Guttorp, P., & Gneiting, T. (2006). Studies in the History of Probability and Statistics XLIX on the Matérn Correlation Family. *Biometrika*, *93*(4), 989–995. Retrieved from http://www.jstor.org/stable/20441340

Helmstetter, A. (2003, 7). Is Earthquake Triggering Driven by Small Earthquakes? *Physical Review Letters*, *91*(5), 058501. Retrieved from https://link.aps.org/doi/10.1103/PhysRevLett.91.058501 doi: 10.1103/PhysRevLett.91.058501

Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research: Solid Earth*, *110*(5), 1–13. Retrieved from http://doi.wiley.com/10.1029/2004JB003286 doi: 10.1029/2004JB003286

Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2007, 1). High-resolution Time-independent Grid-based Forecast for M ¿= 5 Earthquakes in California. *Seismological Research Letters*, *78*(1), 78–86. Retrieved from https://pubs.geoscienceworld.org/srl/article/78/1/78-86/143328http://srl.geoscienceworld.org/cgi/doi/10.1785/gssrl.78.1.78 doi: 10.1785/gssrl.78.1.78

Helmstetter, A., & Sornette, D. (2002a, 12). Diffusion of epicenters of earthquake aftershocks, Omoris law, and generalized continuous-time random walk models. *Physical Review E*, *66*(6), 061104. Retrieved from https://link.aps.org/doi/10.1103/PhysRevE.66.061104 doi: 10.1103/PhysRevE.66.061104

Helmstetter, A., & Sornette, D. (2002b, 10). Subcritical and supercritical regimes in epidemic models of earthquake aftershocks. *Journal of Geophysical Research: Solid Earth*, *107*(B10), 10–1. Retrieved from https://doi.org/10.1029/2001JB001580 doi: 10.1029/2001JB001580

Helmstetter, A., & Sornette, D. (2003, 10). Båth's law derived from the Gutenberg-Richter law and from aftershock properties. *Geophysical Research Letters*, *30*(20). Retrieved from http://doi.wiley.com/10.1029/2003GL018186 doi: 10.1029/2003GL018186

Hijmans, R. J. (2019). *raster: Geographic Data Analysis and Modeling.* Retrieved from https://CRAN.R-project.org/package=raster

Huc, M., & Main, I. G. (2003, 7). Anomalous stress diffusion in earthquake triggering: Correlation length, time dependence, and directionality. *Journal of Geophysical Research: Solid Earth*, *108*(B7). doi: 10.1029/2001jb001645

Illian, J. B., Sørbye, S. H., & Rue, H. (2012, 12). A toolbox for fitting complex spatial point process models using integrated nested Laplace approximation (INLA). *The Annals of Applied Statistics*, *6*(4), 1499–1530. Retrieved

from http://arxiv.org/abs/1301.1817http://dx.doi.org/10.1214/11-AOAS530http://projecteuclid.org/euclid.aoas/1356629049 doi: 10.1214/11-AOAS530

Jordan, T. H., Chen, Y. T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., ... Zschau, J. (2011, 8). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, *54*(4), 319–391. Retrieved from https://www.annalsofgeophysics.eu/index.php/annals/article/view/5350 doi: 10.4401/ag-5350

Jordan, T. H., & Jones, L. M. (2010, 7). Operational Earthquake Forecasting: Some Thoughts on Why and How. *Seismological Research Letters*, *81*(4), 571–574. Retrieved from https://pubs.geoscienceworld.org/srl/article/81/4/571-574/143702 doi: 10.1785/gssrl.81.4.571

Kahle, D., & Wickham, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, *5*(1), 144–161. Retrieved from http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf

Kreemer, C., Blewitt, G., & Klein, E. C. (2014, 10). A geodetic plate motion and Global Strain Rate Model. *Geochemistry, Geophysics, Geosystems*, *15*(10), 3849–3889. Retrieved from http://doi.wiley.com/10.1002/2014GC005407 doi: 10.1002/2014GC005407

Lindgren, F., Rue, H., & Lindström, J. (2011, 9). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(4), 423–498. Retrieved from http://doi.wiley.com/10.1111/j.1467-9868.2011.00777.x doi: 10.1111/j.1467-9868.2011.00777.x

Lindgren, F., & Rue, H. a. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software*, *63*(19), 1–26. Retrieved from http://www.jstatsoft.org/http://www.jstatsoft.org/v63/i19 doi: 10.18637/jss.v063.i19

Lombardo, L., Opitz, T., & Huser, R. (2018). Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. *Stochastic Environmental Research and Risk Assessment*, *32*(7), 2179–2198. Retrieved from https://doi.org/10.1007/s00477-018-1518-0 doi: 10.1007/s00477-018-1518-0

Main, I. G. (1995, 10). Earthquakes as critical phenomena: Implications for probabilistic seismic hazard analysis. *Bulletin of the Seismological Society of America*, *85*(5), 1299–1308.

Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. (2013, 11). Bayesian computing with INLA: New features. *Computational Statistics and Data Analysis*, *67*, 68–83. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167947313001552 doi: 10.1016/j.csda.2013.04.014

Marzocchi, W., Lombardi, A. M., & Casarotti, E. (2014, 9). The Establishment of an Operational Earthquake Forecasting System in Italy. *Seismological Research Letters*, *85*(5), 961–969. Retrieved from https://pubs.geoscienceworld.org/srl/article/85/5/961-969/315409http://srl.geoscienceworld.org/cgi/doi/10.1785/0220130219 doi: 10.1785/0220130219

Marzocchi, W., Zechar, J. D., & Jordan, T. H. (2012, 12). Bayesian Forecast Evaluation and Ensemble Earthquake Forecasting. *Bulletin of the Seismological Society of America*, *102*(6), 2574–2584. Retrieved from https://doi.org/10.1785/0120110327 doi: 10.1785/0120110327

Michael, A. J., & Werner, M. J. (2018, 7). Preface to the Focus Section on the Collaboratory for the Study of Earthquake Predictability (CSEP): New Results and Future Directions. *Seismological Research Letters*, *89*(4), 1226–1228. Retrieved from https://pubs.geoscienceworld.org/ssa/srl/article/89/4/

1226/532049/Preface-to-the-Focus-Section-on-the-Collaboratory doi: 10.1785/0220180161

Murphy, A. H. (1993, 6). What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, *8*(2), 281–293. Retrieved from https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO2 doi: 10.1175/1520-0434(1993)008⟨0281:WIAGFA⟩2.0.CO;2

Nandan, S., Ouillon, G., Wiemer, S., & Sornette, D. (2017, 7). Objective estimation of spatially variable parameters of epidemic type aftershock sequence model: Application to California. *Journal of Geophysical Research: Solid Earth*, *122*(7), 5118–5143. Retrieved from http://doi.wiley.com/10.1002/2016JB013266 doi: 10.1002/2016JB013266

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes.* Retrieved from https://CRAN.R-project.org/package=RColorBrewer

Ogata, Y. (1988, 3). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, *83*(401), 9–27. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/01621459.1988.10478560 doi: 10.1080/01621459.1988.10478560

Ogata, Y. (1998, 6). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, *50*(2), 379–402. Retrieved from http://link.springer.com/10.1023/A:1003403601725 doi: 10.1023/A:1003403601725

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space*, *63*(3), 217–229. Retrieved from http://link.springer.com/10.5047/eps.2010.09.001 doi: 10.5047/eps.2010.09.001

Ogata, Y., & Zhuang, J. (2006, 2). Spacetime ETAS models and an improved extension. *Tectonophysics*, *413*(1-2), 13–23. Retrieved from https://www.sciencedirect.com/science/article/pii/S0040195105004889http://www.sciencedirect.com/science/article/pii/S0040195105004889 doi: 10.1016/J.TECTO.2005.10.016

Pebesma, E. E. J., & Bivand, R. R. S. (2005, 11). Classes and methods for spatial data in R. *R News*, *5*(2), 9–13. Retrieved from https://cran.r-project.org/doc/Rnews/

Rhoades, D. A., Christophersen, A., & Gerstenberger, M. C. (2015, 11). Multiplicative Earthquake Likelihood Models Based on Fault and Earthquake Data. *Bulletin of the Seismological Society of America*, *105*(6), 2955–2968. Retrieved from https://doi.org/10.1785/0120150080 doi: 10.1785/0120150080

Rhoades, D. A., Christophersen, A., Gerstenberger, M. C., Liukis, M., Silva, F., Marzocchi, W., ... Jordan, T. H. (2018, 7). Highlights from the First Ten Years of the New Zealand Earthquake Forecast Testing Center. *Seismological Research Letters*, *89*(4), 1229–1237. Retrieved from https://pubs.geoscienceworld.org/ssa/srl/article/89/4/1229/532037/Highlights-from-the-First-Ten-Years-of-the-New doi: 10.1785/0220180032

Rhoades, D. A., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., Schorlemmer, D., Werner, M. J., & Jordan, T. H. (2014, 9). Regional Earthquake Likelihood Models II: Information Gains of Multiplicative Hybrids. *Bulletin of the Seismological Society of America*, *104*(6), 3072–3083. Retrieved from https://doi.org/10.1785/0120140035 doi: 10.1785/0120140035

Rue, H., Martino, S., & Chopin, N. (2009, 4). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *71*(2), 319–392. Retrieved from http://doi.wiley.com/10.1111/j.1467-9868.2008.00700.x doi: 10.1111/j.1467-9868.2008.00700.x

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., & Lindgren, F. K.

(2017, 3). Bayesian Computing with INLA: A Review. *Annual Review of Statistics and Its Application*, *4*(1), 395–421. Retrieved from http://www.annualreviews.org/doi/10.1146/annurev-statistics-060116-054045 doi: 10.1146/annurev-statistics-060116-054045

Sadykova, D., Scott, B. E., De Dominicis, M., Wakelin, S. L., Sadykov, A., & Wolf, J. (2017, 7). Bayesian joint models with INLA exploring marine mobile predatorprey and competitor species habitat overlap. *Ecology and Evolution*, *7*(14), 5212–5226. Retrieved from http://doi.wiley.com/10.1002/ece3.3081 doi: 10.1002/ece3.3081

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., & Rhoades, D. A. (2007, 1). Earthquake Likelihood Model Testing. *Seismological Research Letters*, *78*(1), 17–29. Retrieved from https://pubs.geoscienceworld.org/srl/article/78/1/17-29/143320 doi: 10.1785/gssrl.78.1.17

Schorlemmer, D., Werner, M. J., Marzocchi, W., Jordan, T. H., Ogata, Y., Jackson, D. D., ... Zhuang, J. (2018, 7). The Collaboratory for the Study of Earthquake Predictability: Achievements and Priorities. *Seismological Research Letters*, *89*(4), 1305–1313. Retrieved from https://pubs.geoscienceworld.org/ssa/srl/article/89/4/1305/532043/The-Collaboratory-for-the-Study-of-Earthquake doi: 10.1785/0220180053

Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017, 1). Estimating ETAS: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469. Retrieved from http://doi.wiley.com/10.1002/2016JB012809 doi: 10.1002/2016JB012809

Simpson, D., Illian, J. B., Lindgren, F., Sørbye, S. H., & Rue, H. (2015, 3). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, *103*(1), 49–70. Retrieved from https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/asv064 doi: 10.1093/biomet/asv064

Simpson, D., Lindgren, F., & Rue, H. (2012, 5). Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, *1*, 16–29. Retrieved from https://www.sciencedirect.com/science/article/pii/S2211675312000048 doi: 10.1016/j.spasta.2012.02.003

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002, 10). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(4), 583–639. Retrieved from http://doi.wiley.com/10.1111/1467-9868.00353 doi: 10.1111/1467-9868.00353

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2014, 6). The deviance information criterion: 12 years on. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(3), 485–493. Retrieved from http://doi.wiley.com/10.1111/rssb.12062 doi: 10.1111/rssb.12062

Strader, A., Werner, M., Bayona, J., Maechling, P., Silva, F., Liukis, M., & Schorlemmer, D. (2018, 7). Prospective Evaluation of Global Earthquake Forecast Models: 2 Yrs of Observations Provide Preliminary Support for Merging Smoothed Seismicity with Geodetic Strain Rates. *Seismological Research Letters*, *89*(4), 1262–1271. Retrieved from https://pubs.geoscienceworld.org/ssa/srl/article/89/4/1262/532041/Prospective-Evaluation-of-Global-Earthquake doi: 10.1785/0220180051

Taroni, M., Marzocchi, W., Schorlemmer, D., Werner, M. J., Wiemer, S., Zechar, J. D., ... Euchner, F. (2018, 7). Prospective CSEP Evaluation of 1Day, 3Month, and 5Yr Earthquake Forecasts for Italy. *Seismological Research Letters*, *89*(4), 1251–1261. Retrieved from https://pubs.geoscienceworld.org/ssa/srl/article/89/4/1251/532038/Prospective-CSEP-Evaluation-of-1Day-3Month-and-5Yr doi: 10.1785/0220180031

Taylor, B. M., & Diggle, P. J. (2014, 10). INLA or MCMC? A tutorial and comparative evaluation for spatial prediction in log-Gaussian Cox processes. *Journal of Statistical Computation and Simulation*, *84*(10), 2266–2284. Retrieved from https://doi.org/10.1080/00949655.2013.788653 doi: 10.1080/00949655.2013.788653

Teng, M., Nathoo, F., & Johnson, T. D. (2017, 7). Bayesian computation for Log-Gaussian Cox processes: a comparative analysis of methods. *Journal of Statistical Computation and Simulation*, *87*(11), 2227–2252. Retrieved from https://doi.org/10.1080/00949655.2017.1326117 doi: 10.1080/00949655.2017.1326117

Tennekes, M. (2018, 4). tmap : Thematic Maps in R. *Journal of Statistical Software*, *84*(6), 1–39. Retrieved from http://www.jstatsoft.org/v84/i06/ doi: 10.18637/jss.v084.i06

Veen, A., & Schoenberg, F. P. (2008, 6). Estimation of SpaceTime Branching Process Models in Seismology Using an EMType Algorithm. *Journal of the American Statistical Association*, *103*(482), 614–624. Retrieved from https://doi.org/10.1198/016214508000000148 doi: 10.1198/016214508000000148

Vere-Jones, D. (1970, 4). Stochastic Models for Earthquake Occurrence. *Journal of the Royal Statistical Society. Series B (Methodological)*, *32*(2), 1–62. Retrieved from https://academic.oup.com/gji/article-lookup/doi/10.1111/j.1365-246X.1975.tb05893.xhttps://www.jstor.org/stable/2984402 doi: 10.2307/2984402

Vere-Jones, D. (1995). Forecasting earthquakes and earthquake risk. *International Journal of Forecasting*, *11*(4), 503–538. Retrieved from http://www.forecastingprinciples.com/paperpdf/Vere-Jones-forecastingearthquakes.pdf doi: 10.1016/0169-2070(95)00621-4

Vere-Jones, D., & Davies, R. B. (1966, 10). A statistical survey of earthquakes in the main seismic region of New Zealand. *New Zealand Journal of Geology and Geophysics*, *9*(3), 251–284. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00288306.1966.10422815 doi: 10.1080/00288306.1966.10422815

Werner, M. J., Helmstetter, A., Jackson, D. D., & Kagan, Y. Y. (2011, 8). High-Resolution Long-Term and Short-Term Earthquake Forecasts for California. *Bulletin of the Seismological Society of America*, *101*(4), 1630–1648. Retrieved from https://pubs.geoscienceworld.org/bssa/article/101/4/1630-1648/349546 doi: 10.1785/0120090340

Yuan, Y., Bachl, F. E., Lindgren, F., Borchers, D. L., Illian, J. B., Buckland, S. T., ... Gerrodette, T. (2017, 12). Point process models for spatio-temporal distance sampling data from a large-scale survey of blue whales. *Annals of Applied Statistics*, *11*(4), 2270–2297. Retrieved from https://projecteuclid.org/euclid.aoas/1514430286 doi: 10.1214/17-AOAS1078

Zechar, D. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H. (2013, 4). Regional Earthquake Likelihood Models I: First-order results. *Bulletin of the Seismological Society of America*, *103*(2A), 787–798. Retrieved from https://pubs.geoscienceworld.org/bssa/article/103/2A/787-798/331692 doi: 10.1785/0120120186

Zeng, Y., & Shen, Z. (2016, 4). A FaultBased Model for Crustal Deformation, Fault Slip Rates, and OffFault Strain Rate in California. *Bulletin of the Seismological Society of America*, *106*(2), 766–784. Retrieved from https://pubs.geoscienceworld.org/bssa/article/106/2/766-784/332507 doi: 10.1785/0120140250